Summary. Lagrangian analysis of how the northward extension of the North Atlantic Current (across the Eastern half of the OSNAP East section) contributes to the formation of North Atlantic Deep Waters by exposing light "upper limb" waters to buoyancy loss due to surface buoyancy fluxes or turbulent entrainment into denser water masses. Variability in the (Lagrangian) North Atlantic Deep Water formation rate is attributed to the overall strength of the Subpolar Gyre, which is connected to the water mass properties of the gyre's interior (via an assumed thermal wind relationship). These results are contrasted with the idea that the advection of buoyancy anomalies controls the formation of deep waters, as in the salt-advection feedback thought to control the overall stability of the AMOC.

Overall Assessment. This is an excellent manuscript. The Lagrangian diapycnal overturning framework provides some unique insights into the overturning of the Subpolar North Atlantic. The authors also take this further by introducing the concept of a transformation efficiency, which they show to be useful because it is constant at leading order, meaning that the initial transport of the NAC is the main driver of variability. A common problem with new approaches like this is that they are difficult to compare with other studies in the literature. However, the authors successfully connect (or contrast) their results to prior results–most of which are Eulerian.

I recommend this manuscript accepted in Ocean Science after the authors address the minor comments below.

Key Minor Comments:

- In Figure 7 and L. 328-332: As far as I can tell, there is nothing from the analysis in this paper that attribute the diapycnal transformations to specific processes. More specifically, I don't see how you can infer that the RT waters are only mixing with ISOW and not also transformed by surface buoyancy fluxes. If this is an inferrence/speculation baked on prior analysis in Tooth et al. (2023) or other unrelated papers, then you should make that explicit.
- I find it confusing to call your main metric a "Lagrangian diapycnal overturning streamfunction" when it is not actually a streamfunction in the typical sense, but instead is just a difference between two Lagrangian transports. I find it much more appropriate to call it a "Lagrangian diapycnal transformation rate", which you do in some places.
- The specific conclusion that thermohaline anomalies are decoupled from diapycnal overturning seems to me to be a bit overblown. It is not obvious that the same result would apply to the Eulerian diapycnal overturning, which is what is measured by OSNAP and considered in conceptual models of the salt-advection feedback and AMOC stability.
- I encourage the authors to add more discussion of their Lagrangian experiment design, its caveats, and why they picked it over alternative approaches. If the goal was instead to understand the variability of NADW transported southwards across OSNAP West, then a

backtracking experiment would have been more appropriate, whereby particles are grouped according to their final time and convolved over many different release times. Would the authors expect to get the same qualitative results in that case?

Minor line-by-line comments:

- The acronym NAC is never defined!
- Can you explain why you use potential density referenced to the surface for the Lagrangian analysis but referenced to 2000 dbar for the Eulerian water mass analyses?
- Equation (4) seems incorrect to me. First, shouldn't the sum be over all density layers, since the box function is already picking out just the discrete outcropping layer? Second, I think you need to divide by the size of the density bin?
- L. 200-210 and Figure 3. Upon first read it is really hard to keep track of all of these transport numbers and how they are related. This is made even more difficult because the way you've rounded numbers means that things don't add up in a consistent way. For example, I was confused why the transports of 21.2 Sv and 3.9 Sv in Figure 3(b) did not add up to the total NAC transport 24.8 Sv. Can you round these up or down so that they're all self-consistent?
- Figure 3 panel labels are inconsistent with the description in the caption.
- Figure 3: Can you add the time-mean isopycnal that separates the two branches, \sigma_{DWF}?
- L. 208-210: Explain this comparison with observations better. Are the first two references some kind of analogous Lagrangian estimate of transports? Or an Eulerian transport but just for the strictly northward transport into the Nordic Seas whereas the 5.8 Sv (Østerhus) estimate is for the total transport?
- L. 222-225: The phrasing here is a bit confusing, especially because of the first sentence. I think what you mean is that: "Because 5.6 Sv of the water flowing northwards across OSNAP East is already in the lower limb, the 12.7 Sv of NADW formation is in fact a relatively larger fraction of the 19.2 Sv that is in the upper limb."
- L. 228: Can you cite a specific result from OSNAP here?
- L. 263: Worth emphasizing here (and perhaps in other places where it may be ambiguous) that the time t always refers to the time of "release", not the time at which transformation actually occurs or when it leaves across OSNAP West.
- L. 345 and Figure 8b-c. This is not a very interesting result and I think is mostly explained by the application of a binary sorting based on a fixed density threshold. Of course waters with inflow densities much less than \sigma_{DWF} will need to transform more in order to cross the threshold. I would just cut these two panels.
- Figure 9b. What is the point of showing such a broad range of densities when we're only meant to focus on \sigma_{2}=37.0? Can you either plot this as a timeseries or zoom in on the denser waters a bit?

- Figure 9b Caption: Clarify that these are (I assume) anomalies relative to a monthly climatology.
- L. 403-408: This is really difficult to parse as written. I think what you mean is that, because dense waters do not outcrop in the western Labrador Sea, that suggests there is no significant formation of local NADW from waters coming north across the western part of OSNAP-W. But I don’t really understand how that implies that \kappa < 100%.
- L. 479: “neither the efficiency of along-stream diapycnal transformation”...